PNAS
www.pnas.org

Supplementary Information for

Stochastic microbiome assembly depends on context

Eric W. Jones, Jean M. Carlson, David A. Sivak, and William B. Ludington

Eric W. Jones
E-mail: eric_jones_2@sfu.ca

This PDF file includes:

Supporting Information Text

Results

Clustering the dendrogram of Fig. 3d. A similarity matrix was created with entries equal to the distance between each pair of rows of the heatmap in Fig. 3d, where the distance metric is the 2-norm of the vector of the differences of the three comparable entries (the diagonal elements of every row are ill-defined and cannot be compared). The five rows of the heatmap then were hierarchically clustered based on this similarity matrix using the `scipy.cluster.hierarchy.linkage` Python function, and the results of this clustering were plotted as a dendrogram with the `scipy.cluster.hierarchy.dendrogram` Python function.

Replacing 100% colonization odds by 99.99% versus 99%. Any independent colonization model in which the colonization probability of a species is 100% will predict that likelihood of that species not colonizing is precisely 0. This model would subsequently make poor predictions for our system, since in our experiments it was common for even the strongest colonizers to fail to colonize some of the time. To avoid these divergences in the log-likelihood, we replaced the colonization odds of species that colonized 100% of the time by 99.99%. The only model affected by this modification is the single-species model, in which LP, LB, and AO all colonized 100% of the time. Since the choice of the upper bound was arbitrary, we also examined the performance of the single-species model when the upper bound was set to 99%. Whether the upper bound was set to 99% or 99.99%, the single-species model is substantially outperformed (in terms of number of combinations reproduced, log-likelihood, and BIC scores) by the two-species, max-likelihood, and both interaction models. Specifically, the 99% model reproduces 10 of the 31 combinations ($p > 0.05$, multinomial test), has a log-likelihood of -483, and has a BIC score of 965, while the 99.99% model reproduces 8 of the 31 combinations, has a log-likelihood of -1023, and has a BIC score of 2046.

Parameter estimates for the interaction model. The interaction parameters α_{LL}, α_{LA}, α_{AL}, and α_{AA} encode interactions between *Lactobacillus* and *Acetobacter* species. These parameter values were determined by maximizing the likelihood across all experiments, which yielded $\alpha_{LL} = 2.4$, $\alpha_{LA} = 2.0$, $\alpha_{AA} = 2.6$, and $\alpha_{AL} = 0.5$. Therefore, in the interaction model the colonization probabilities of *Lactobacillus* species are lower in the presence of any additional species. The colonization probabilities of *Acetobacter* species are lower in the presence of other *Acetobacter* species and are higher in the presence of *Lactobacillus* species. A given interaction strength α will influence the colonization probabilities of moderate colonizers (*i.e.,* those that colonize $20\% - 60\%$ of the time) more than strong colonizers due to the functional form of $|p_i - p_i^\alpha|$, which has a single maximum at $p_i = \alpha^{1/(1-\alpha)}$. For the interaction parameters we consider ($0.5 < \alpha < 2.6$), $|p_i - p_i^\alpha|$ is greatest when p_i is between 0.2 and 0.6, and decreases as p_i increases further. For example, in the interaction model the colonization probability of AP (the weakest colonizer) decreases by 30 percentage points when it is in the presence of other *Acetobacter* species, yet the colonization probability of LP (the strongest colonizer) decreases by only 2 percentage points when it is in the presence of other *Lactobacillus* species. Colonization probabilities and interaction parameters in the interaction model were fit using the SLSQP method in the `scipy.minimize` Python function.

Leave-k-out cross-validation. To examine whether the max-likelihood and interaction colonization models are robust to subsampling of data, we performed leave-k-out cross validation, in which models are fit with data from $31 - k$ randomly selected combinations and then are tested on their ability to predict the colonization outcomes of the remaining k combinations for $k = 1$ and 15. To generate statistics we repeat this random partitioning process 1,000 times for $k = 15$, and for $k = 1$ we exhaustively consider all 31 training and test sets (1). For each model realization, we computed the log-likelihood of that model generating the training data (composed of $31 - k$ combinations) and generating the test data (composed of k combinations). Fig. S3 shows the average log-likelihood per combination (defined as the log-likelihood over a set of combinations divided by the number of combinations in that set) for both the training and test data. As expected, we find that models perform better (*i.e.,* possess larger average log-likelihoods per combination) on training data than on test data. For smaller k, model performance becomes more similar on training versus test data, with average log-likelihoods varying by less than 15% for both the max-likelihood and interaction models when $k = 1$.

Synthetic data generation and convergence of parameter estimates. To estimate how much data would be required to infer colonization probabilities p_i and deviations $\Delta p^j(i)$ in colonization probabilities of more diverse microbiomes—murine or human gut microbiomes, for example—we generate synthetic colonization data and measure how quickly our estimates $\hat{p}_i$ and $\hat{p}^j(i)$ converge to their ground-truth values p_i and $p^j(i)$. We consider a hypothetical 20-species microbiome and analyze a scenario in which we obtain presence/absence data from colonization experiments of 10-species subsets of the 20-species microbiome.

Synthetic p_i and $\Delta p^j(i)$ parameters are created from bare colonization probabilities $\tilde{p}_i$ (based on the measured colonization odds of the five core bacterial species discussed in the main text) and interaction parameters $\tilde{\alpha}_{ij}$ (which dictate how the colonization of species i is affected by the presence of species j). The bare colonization probabilities $\tilde{p}_i$ are generated for $i \in 1, \ldots, 20$ by taking random convex combinations of the measured colonization odds of the five core *Acetobacter* and *Lactobacillus* species discussed in the main text (*i.e.,* for each $i \in 1, \ldots, 20$, $\tilde{p}_i = \mathbf{R} \cdot \mathbf{P}$ where $\mathbf{R}$ is a five-element vector of random numbers between 0 and 1 with unit sum and $\mathbf{P}$ is the vector of observed colonization odds of LP, LB, AP, AT, and AO). Figure S4 shows these bare colonization probabilities $\tilde{p}_i$ as red dots.

To generate interactions, we assume that this 20-species microbiome is composed of four five-species genera (labeled A, B, C, and D), and we further assume that the colonization probabilities are affected by interactions such that, assuming species i

is in genus A, its colonization probability is

$$p_i = \tilde{p}_i^{\,\tilde{\alpha}_{\mathrm{AA}}^{(n_{\mathrm{A}}-1)}\tilde{\alpha}_{\mathrm{AB}}^{n_{\mathrm{B}}}\tilde{\alpha}_{\mathrm{AC}}^{n_{\mathrm{C}}}\tilde{\alpha}_{\mathrm{AD}}^{n_{\mathrm{D}}}}, \quad [1]$$

where n_J indicates the number of species of the 10-species microbiome from genus J. p_i is similarly defined when $i \in$ B, C, or D. We choose the interaction matrix

$$\tilde{\alpha} = \begin{bmatrix} 1.3 & 1 & 1 & 1 \\ 0.7 & 1.3 & 1 & 1.3 \\ 1 & 1.3 & 1.3 & 1.3 \\ 1.3 & 1 & 1 & 1 \end{bmatrix}, \quad [2]$$

where each element $\tilde{\alpha}_{ij}$ is given by the ith row and jth column, and where A, B, C, and D correspond to i or $j = 1$, 2, 3, and 4, respectively. In all, the bare colonization probabilities $\tilde{p}_i$ and interactions $\tilde{\alpha}$ are used to generate synthetic colonization data from which colonization probabilities p_i (black dots in Fig. S4) and deviations $\Delta p^j(i)$ in colonization probabilities (heatmap in Fig. S4) may be inferred.

Figure S5 shows the expectation of the average absolute deviations $\mathbb{E}\left[\frac{1}{M}\sum_i |\hat{p}_i - p_i|\right]$ and $\mathbb{E}\left[\frac{1}{M^2}\sum_{i,j} |\hat{p}^j(i) - p^j(i)|\right]$ as a function of the number N of 10-species combinations tested. These expectation values describe on average how close each colonization probability or deviation in colonization probability is to its ground-truth value, and the expectation is taken over 100 realizations of N 10-species experiments. We consider scenarios in which each experiment contains 12, 24, 48, or 96 replicates. The quantity $\mathbb{E}[|\hat{p}_i - p_i|]$ scales as the standard deviation of $\hat{p}_i - p_i$ (assuming the distribution of $\hat{p}_i - p_i$ is approximately normal (2)), which scales as $1/\sqrt{Q_1}$ where Q_1 is the number of trials in which species i is fed. Similarly, $|\Delta\hat{p}^j(i) - \Delta p^j(i)|$ scales as $1/\sqrt{Q_2}$ where Q_2 is the number of trials in which both species i and species j are fed. On average, for N random 10-species combinations in a 20-species microbiome, $Q_1 = N/2$ and $Q_2 = N/4$. Therefore, the average absolute deviations in the colonization probabilities and the deviations in colonization probabilities both scale as $N^{-1/2}$, as illustrated by the black line in Fig. S5.

Materials and Methods

Pooling colonization data from biological replicates. To determine whether there was any systematic variation between any of the biological replicates, we investigated whether the outcomes observed for each 12-fly replicate are consistent with the outcomes of the corresponding aggregate 48-fly experiment. We used a multinomial test to calculate the probability of obtaining the colonization outcomes of each 12-fly replicate with a null multinomial distribution characterized by the empirical colonization outcomes of the corresponding 48-fly experiment. Figure S6 summarizes the results of these multinomial tests, showing for each of the 4 biological replicates the p-values of generating the colonization outcomes of the 31 combinations. Based on this analysis we find that the data are slightly skewed, with multinomial tests only indicating significant deviations from the null multinomial distribution for experimental combinations AT/AO in replicate 1 and AP/AT/AO in replicate 2. Despite the significant deviation of these two 12-fly replicates, to obtain greater statistical power we use pooled colonization data from all replicates and did not distinguish between replicates in our analyses.

Scaling of confidence intervals with increasing sample size. The mean size of the 95% confidence intervals in Fig. 2 of the main text is 21 percentage points. When the number of samples N is sufficiently large to approximate a binomial distribution by a normal distribution, the standard deviation scales as $N^{-1/2}$. Roughly, this implies that performing 100 times as many experiments (*i.e.,* sampling 4800 flies per combination) would reduce confidence intervals 10-fold (giving a mean confidence interval width of 2 percentage points).

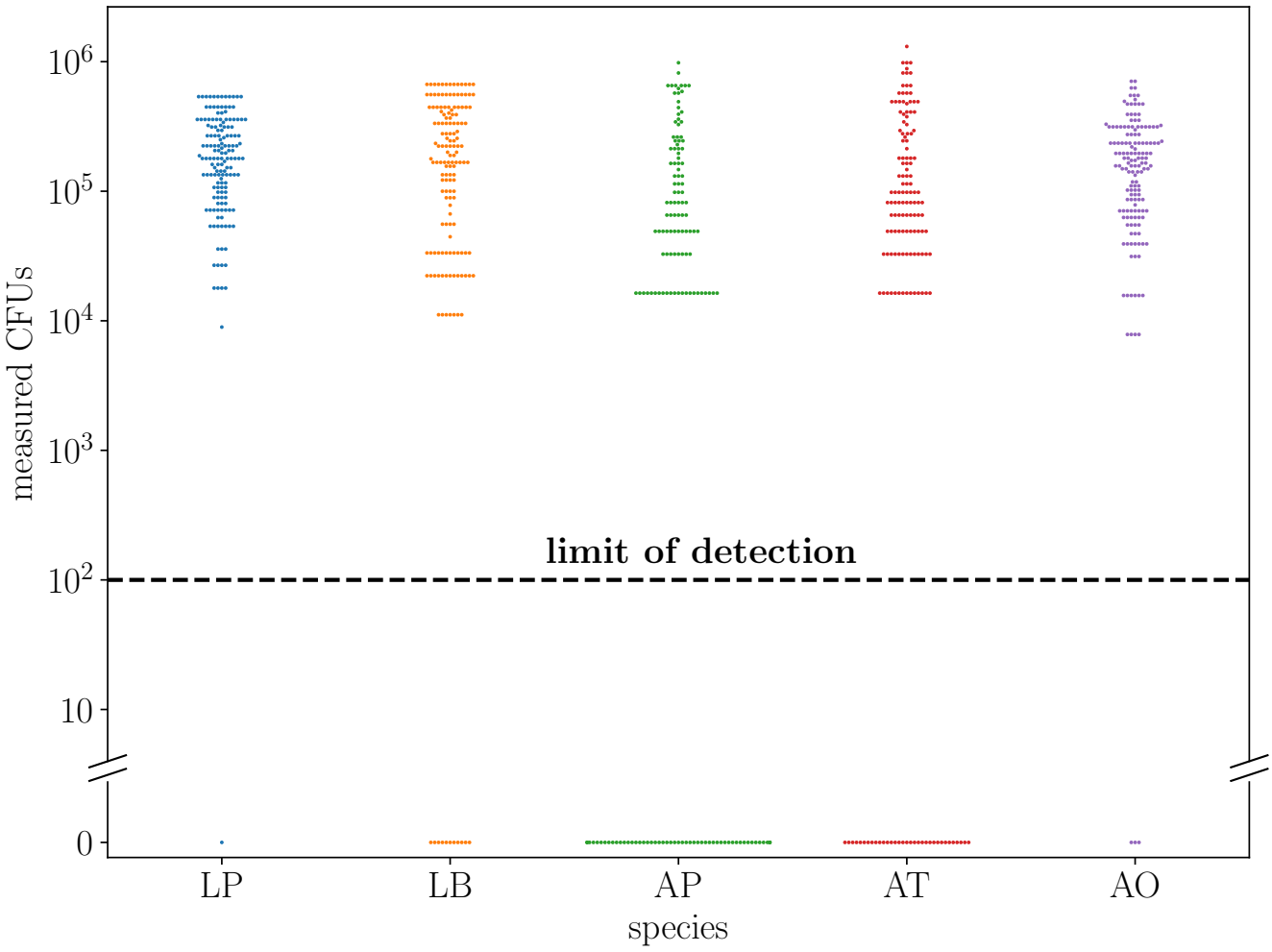

Fig. S1. Successfully colonized bacteria colonize at an abundance substantially higher than the experimental limit of detection. Each data point is the bacterial abundance (measured in colony-forming units (CFUs)) of a biological replicate of a bacterial combination. When bacteria fail to colonize a fly (beyond the detection limit), their abundance is marked zero. A random subset of 10% of data points are plotted.

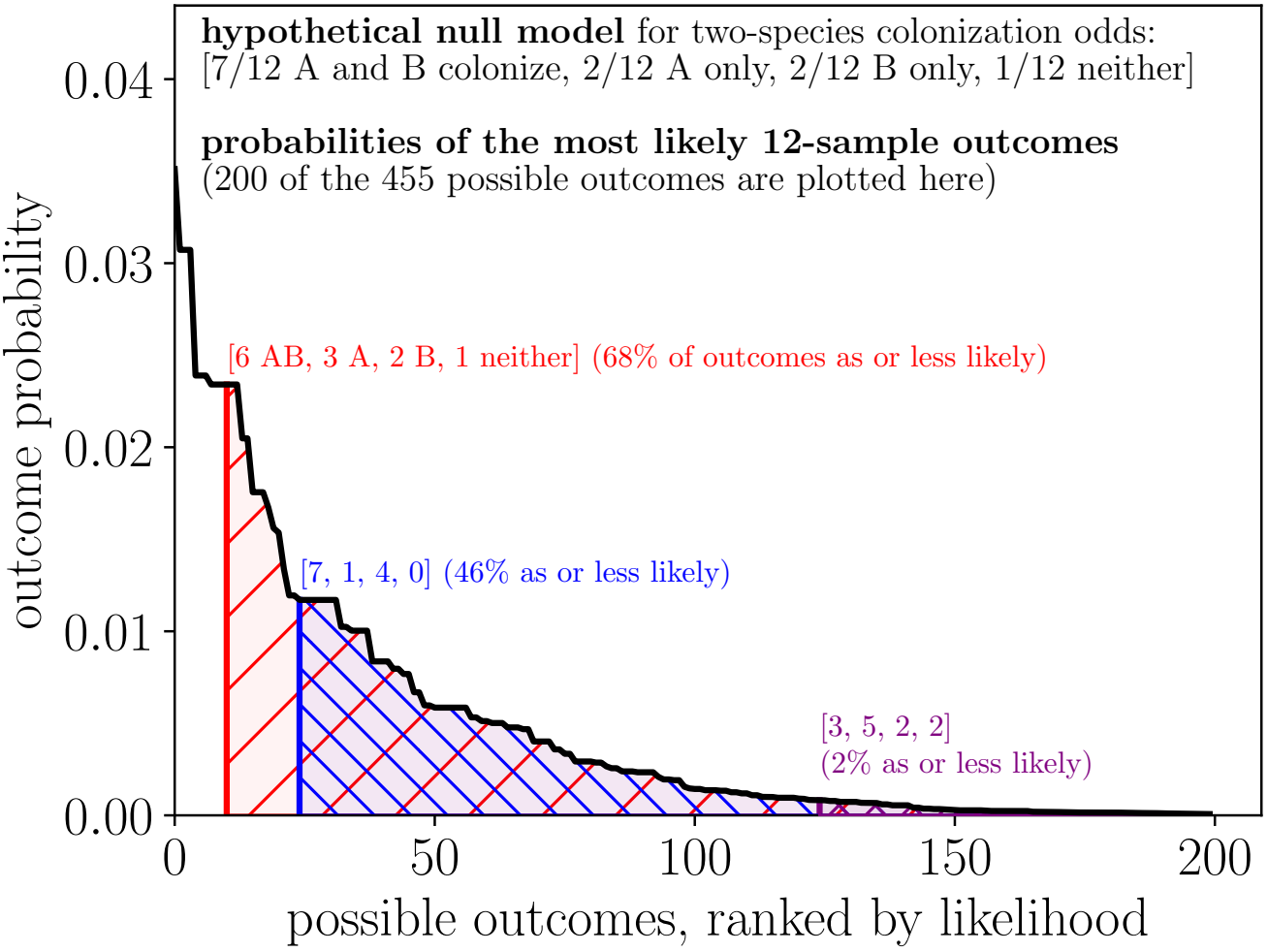

Fig. S2. Explanation of the multinomial test, which tests whether an observed outcome is likely, assuming a null multinomial distribution. In the main text we test six null models, which each predict the distribution of colonization outcomes for each combination of bacteria, and are tested against the experimentally measured outcomes (48 flies per bacterial combination). To evaluate the likelihood of a particular N-element outcome with the exact multinomial test, compute the probability of all possible N-element outcomes, and sort the outcomes according to their probabilities. Then, sum the probability of the outcome of interest and the probabilities of all of the less-likely outcomes (i.e., integrate from the outcome of interest to the right in the figure). This integrated area is the p-value, which is how likely an observed outcome (or any less likely outcome) is under a particular null model.

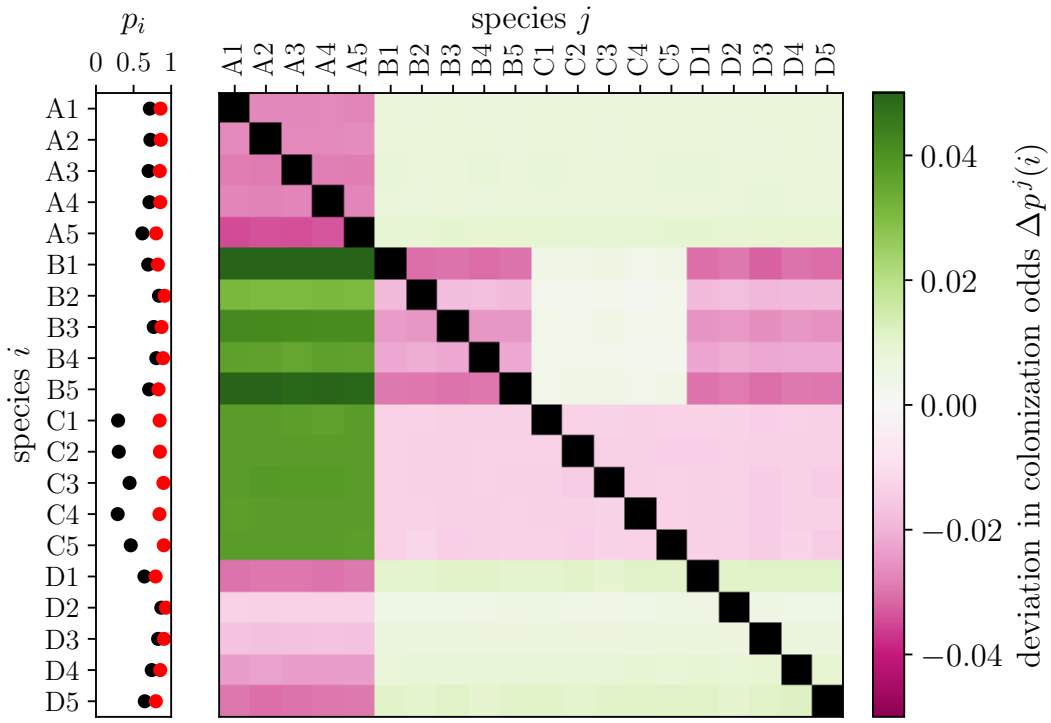

Fig. S3. Synthetically generated colonization probabilities p_i and deviations $\Delta p^j(i)$ in colonization probabilities. Red dots: randomly generated bare colonization probabilities $\tilde{p}_i$. Black dots: colonization probabilities p_i inferred from data from numerous 10-species microbiomes (*i.e.,* probabilities that have been influenced by interactions between species). The ground-truth quantities p_i (black dots) and $\Delta p^j(i)$ are inferred from 200 synthetic experiments of 10-species microbiomes in which each experiment consists of 500 replicates.

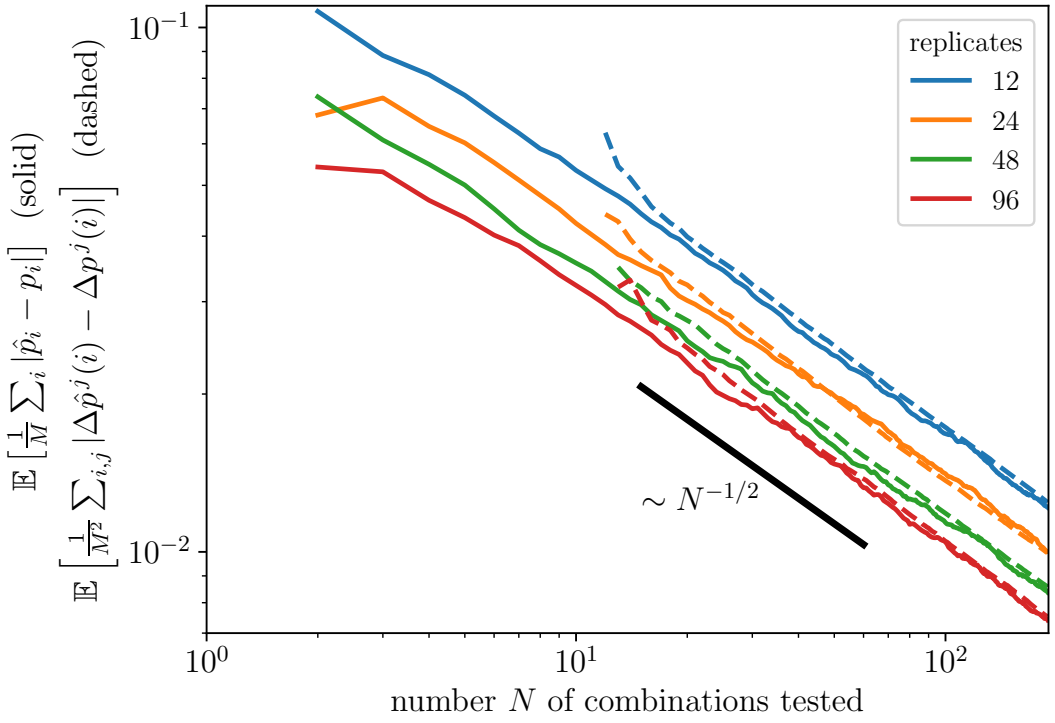

Fig. S4. Estimation of colonization probabilities p_i and deviations $\Delta p^j(i)$ in colonization probabilities as a function of the number N of 10-species combinations tested for a synthetic 20-species microbiome. For each number N of combinations tested, we compute the respective average absolute deviations $\frac{1}{M}\sum_i |\hat{p}_i - p_i|$ (solid curves) and $\frac{1}{M^2}\sum_{i,j} |\Delta\hat{p}^j(i) - \Delta p^j(i)|$ (dashed curves) of estimators $\hat{p}_i$ and $\Delta\hat{p}^j(i)$ inferred from N 10-species combinations and ground-truth quantities p_i and $\Delta p^j(i)$. We estimate the expected values of these average absolute deviations by averaging over 100 synthetic experiments of N 10-species microbiomes, in which each experiment consists of the number of replicates indicated by the legend.

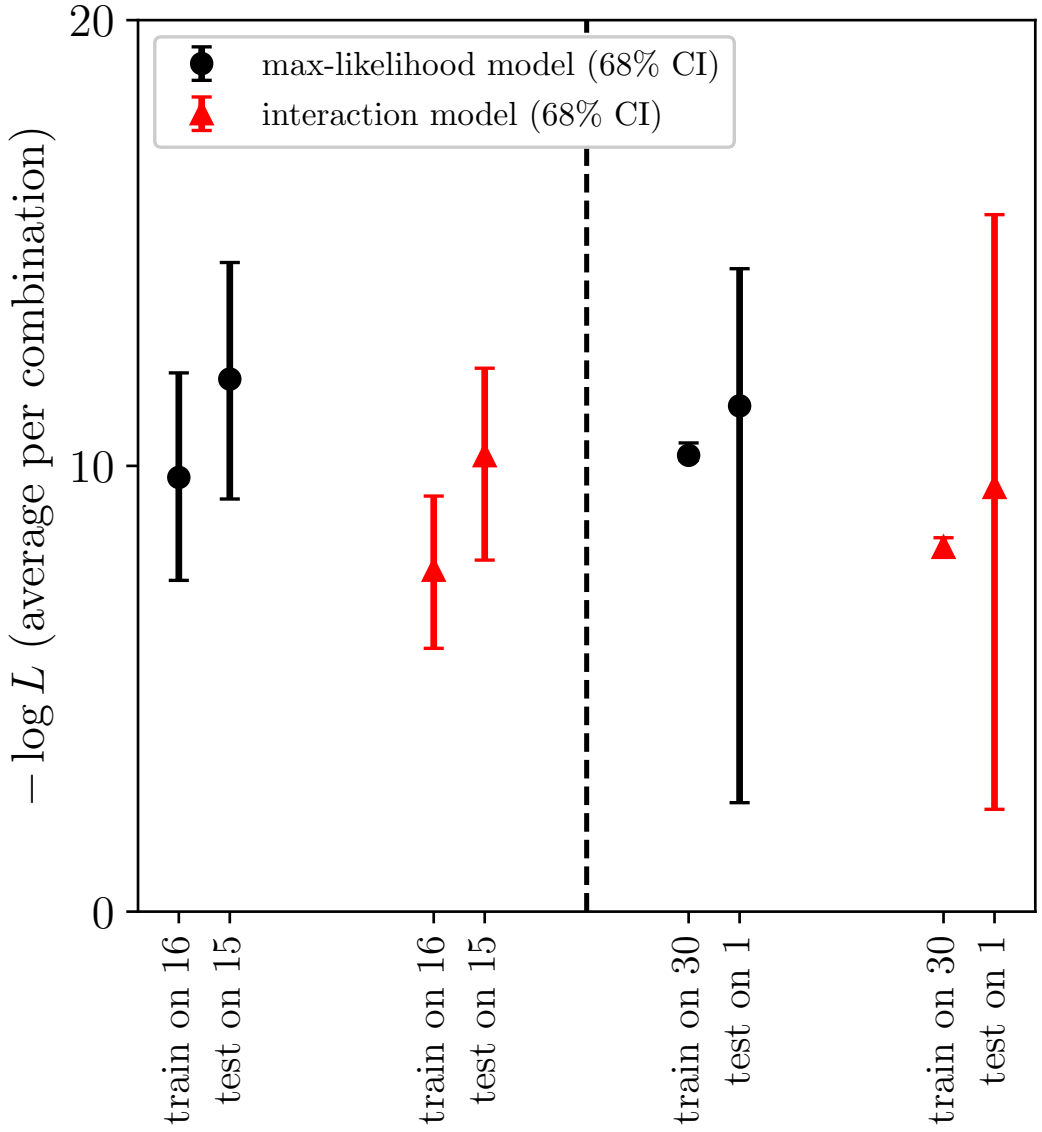

Fig. S5. Leave-k-out cross validation for k=1 and 15. The max-likelihood (black circles) and interaction (red triangles) models are fit to training data from $31 - k$ of the combination experiments, then evaluated on their ability to predict the outcomes of these $31 - k$ combinations (train data, left) and the remaining k combinations (test data, right). Symbols indicate averages and error bars indicate one standard deviation (from complete sampling for $k = 1$, and 1,000 random train/test partitions for $k = 15$).

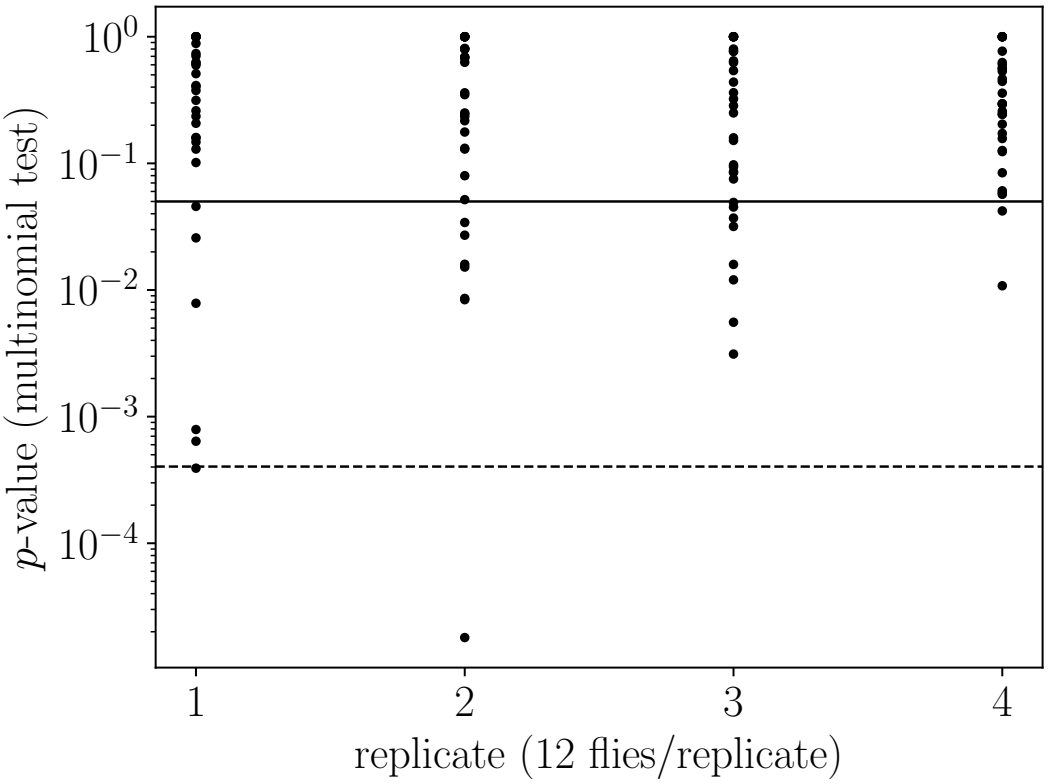

Fig. S6. Multinomial tests across replicates for each combination of fed species. For each of the 31 combinations of fed species, the probability of observing the colonization outcome of each 12-fly replicate using a null multinomial distribution characterized by the pooled 48-fly data is plotted. Solid horizontal line: $p = 0.05$. Dashed line: Bonferroni-corrected significance level of $p = 0.05/(31 * 4) = 4 * 10^{-4}$ which accounts for multiple tests (3).

Table S1. Log-likelihood of independent colonization models for experiments of each diversity.

independent colonization model	log-likelihood (diversity-1 experiments)	log-likelihood (diversity-2 experiments)	log-likelihood (diversity-3 experiments)	log-likelihood (diversity-4 experiments)	log-likelihood (diversity-5 experiments)
uniform	-31	-136	-228	-151	-36
single-species	-4	-201	-536	-202	-79
two-species	-10	-73	-164	-67	-15
max-likelihood	-13	-76	-148	-67	-14
interaction (div-2)	-11	-49	-138	-78	-17
interaction (all)	-8	-54	-109	-70	-13

References

1. CFJ Wu, Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals Stat.* **14**, 1261–1295 (1986).
2. RC Geary, The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika* **27**, 310–332 (1935).
3. OJ Dunn, Multiple comparisons among means. *J. Am. Stat. Assoc.* **56**, 52–64 (1961).